Structural Causal Models and Abstraction for Modeling Battery Manufacturing

Shared work of F.M. Zennaro, M. Drávucz, G. Apachitei,
W.D. Widanage and T. Damoulas

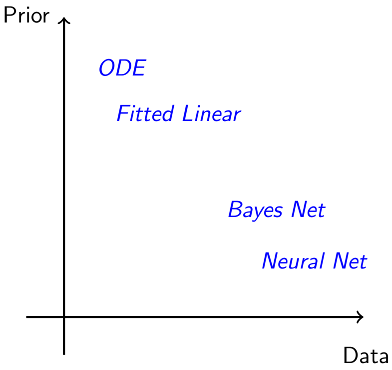
Outline

2. Structural Causal Modelling

Modelling

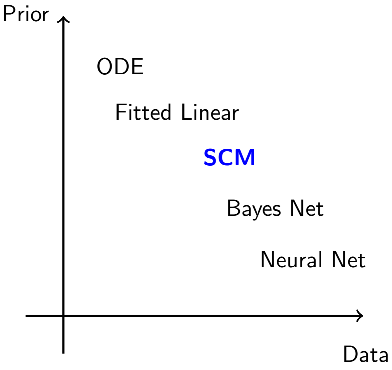
Assume we want to model a *system*.

Different types of model will negotiate a trade-off between priors and data:



Structural Causal Models

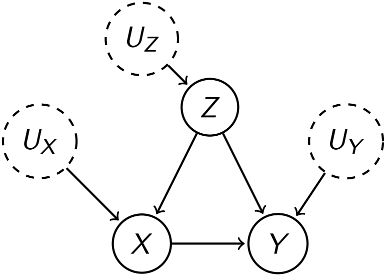
Structural causal models rely on a strong prior given by *causality* [5].



- It discriminates *correlations* and *causes*.
- It allows for reasoning about *interventions*.
- It allows for reasoning about *counterfactuals*.
- It implies a *causality ladder* of reasoning.
- It requires more than data.

Structural causal model

A **SCM** [4, 5] is a mathematical object $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ with an associated DAG:

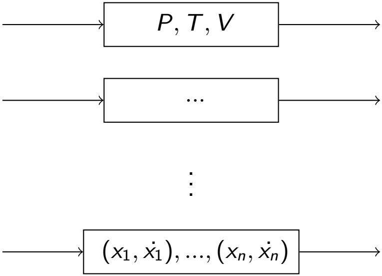


- It adds *structure* to a purely statistical model.
- It does not require to be *fully defined*.
- It allows for *causal inference* [4, 8] and *causal discovery* [3] relying on data.

3. Abstraction

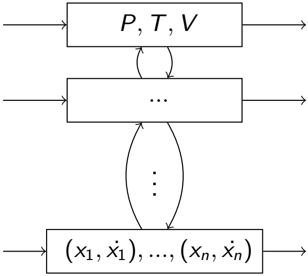
Modelling

A system can be modelled on multiple *levels of abstraction*.



Abstraction

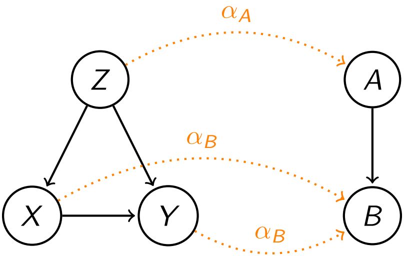
Abstraction (aka, *multi-level modelling* or *multi-resolution modelling*) aims at relating these levels.



- It combines models from *different sources*.
- It aggregates information from *different resolutions*.
- It allows for *computation with minimal effort*.

Causal Abstraction

Causal abstraction is a collection of maps α between two SCMs $\mathcal{M}$ and $\mathcal{M}'$ that relates variables [7, 1, 6].



- It captures *coarsening/change of resolution/simplification of structure*.

Causal Consistency

We impose on the *causal abstraction* a requirement of *interventional consistency* (or *approximate interventional consistency*) [7, 1, 6].

$$\begin{tikzcd} \mathcal{M} \arrow[r, "\alpha"] \arrow[d, "\iota"'] & \mathcal{M}' \arrow[d, "\iota'"] \\ \mathcal{M}_\iota \arrow[r, "\alpha"] & \mathcal{M}'_{\iota'} \end{tikzcd}$$

- Consistency would allow to *switch*/*intergrate* models.
- We have a *formal definition*, limited methodology.

Recap

With **causality** we can:

- Exploit (partial) prior causal knowledge to define causal DAG
- Exploit (partially defined) DAGs to perform causal inference

With **causal abstraction** we can:

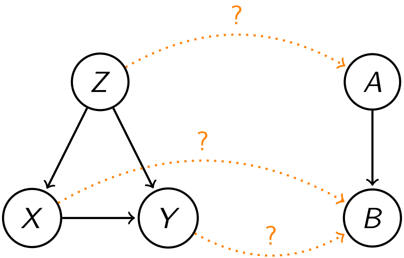
- Exploit causal models at multiple levels to leverage heterogeneous data

- *How do we learn an abstraction if we are not given one?*
- *How do we apply it to the relevant problem of battery manufacturing?*

4. Learning Causal Abstractions

Learning Causal Abstraction

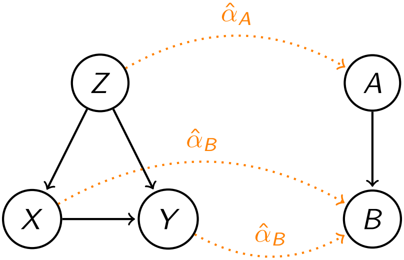
How do we learn the abstraction maps?



- Not trivial: we have to deal with a *combinatorial optimization* problem for each map α_i.
- Optimization of the maps with respect to global consistency introduces *dependencies*.

Jointly Learning Causal Abstraction

We have proposed a first methodology to learn an abstraction.



- It relies on *relaxing* the problem and using *differentiable programming* to learn all maps at once.
- *Heuristic* - highly efficient at the cost of sub-optimality.
- It requires substantial *prior* knowledge (SCM).

5. Modelling Battery Manufacturing

Problem Definition

We want to model the stage of **coating** in lithium-ion battery manufacturing:

$$\text{Mass Loading} = \text{f(input)}$$

Experiments are costly, so we want to integrate data[1] collected by two groups running similar (but not identical) experiments:

LRCS (France)	**WMG** (UK)
Collection of few statistics in each a few stages of battery manufacturing [2].	Collection of detailed space- and time-dependent measurements during coating.

[1] `https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/batt.201900135`
`https://github.com/mattdravucz/jointly-learning-causal-abstraction/`

Datasets

LRCS

- *Input params:* AM composition, S-to-L ratio, comma gap, viscosity.
- *Output params:* mass loading, porosity
- 656 datapoints in 82 configurations.

high-level, wide

WMG

- *Experimental params:* AM composition, ...
- *Machine params:* Comma bar operator position actual, Coating roll gear ratio setpoint, ... measured every 1s
- *Output params:* Mass loading ... measured every 8s at 800 spatial locations
- 1 experiment lasting 3h with varying configurations

low-level, narrow

Preprocessing and alignment

LRCS

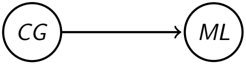
- Unit conversion
  Aligning unit of measure
- Params subselection
  Dropping slurry params
- Discretization
  Binning ML into n_{bins}

WMG

- Params combination
  Reconstructing actual comma gap
- Time subselection
  Filtering transitions and downtimes
- Space averaging
  From 800 locations to n_{loc}
- Discretization
  Binning ML into n_{bins}
- Data Extraploation
  Additional CG values via GPs

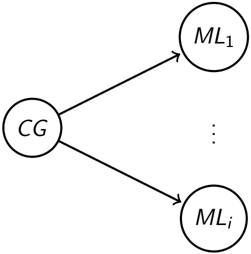
SCM definition

LRCS



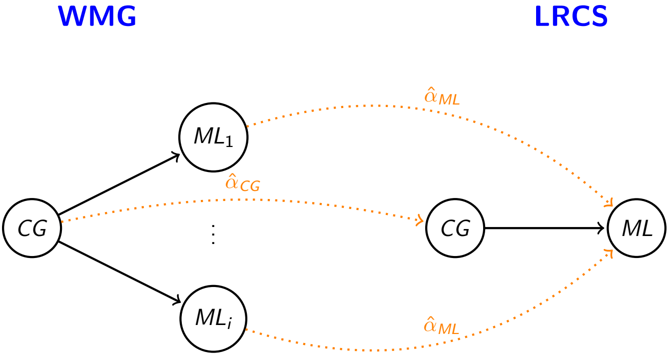
$$\mathcal{M}^{LRCS}[CG] = \{75, 100, 200\}$$
$$\mathcal{M}^{LRCS}[ML] = \{0, 1, ..., n_{bins}\}$$

WMG



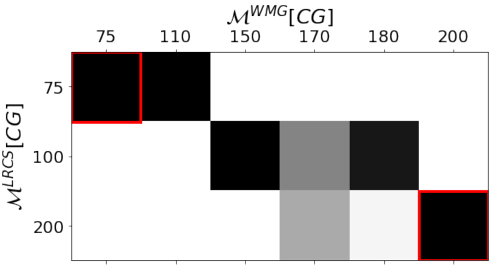
$$\mathcal{M}^{WMG}[CG] = \{75, 110, 150, 170, 180, 200\}$$
$$\mathcal{M}^{WMG}[ML_i] = \{0, 1, ..., n_{bins}\}$$

Learning abstraction



- *Jointly learning causal abstraction*
- Explored a small number of hyperparameters (T, η, λ) across 50 instances.

Results: Insight



- Learning achieves meaningful results *based only on probabilistic and interventional behaviour* (with no reference to semantics).
- Variability remains a challenge

Results: Downstream task

How does aggregation of data really perform on actual problems?

Out-of-sample prediction: LRCS use a LASSO model to predict ML as a function of $CG = k$, when $CG = k$ is far from the training set.

	Training set	**Test Set**	**MSE**
(a)	LRCS[$CG \neq k$]	LRCS[$CG = k$]	1.86 ± 1.75
(b)	LRCS[$CG \neq k$] + WMG	LRCS[$CG = k$]	0.22 ± 0.26
(c)	LRCS[$CG \neq k$] + WMG[$CG \neq k$]	LRCS[$CG = k$] + WMG[$CG = k$]	1.22 ± 0.95

- Transferring data can increase statistical power

Conclusion

- *Causality* and *abstraction* may both play important role in modelling.
- This is first proposal for *learning abstraction*.
- Preliminary results show promise for *combining* diverse data and take advantage of it.

Thanks!

Thank you for listening!

References I

[1] Sander Beckers and Joseph Y Halpern. Abstracting causal models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2678–2685, 2019.

[2] Ricardo Pinto Cunha, Teo Lombardo, Emiliano N Primo, and Alejandro A Franco. Artificial intelligence investigation of nmc cathode manufacturing parameters interdependencies. *Batteries & Supercaps*, 3(1):60–67, 2020.

[3] Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524, 2019.

[4] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[5] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: Foundations and learning algorithms*. MIT Press, 2017.

References II

[6] Eigil Fjeldgren Rischel. The category theory of causal models. 2020.

[7] Paul K Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *33rd Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, pages 808–817. Curran Associates, Inc., 2017.

[8] Jin Tian and Judea Pearl. A general identification condition for causal effects. In *AAAI*, pages 567–573, 2002.